Dynamic Attack Detection in Cyber-Physical Systems with Side Initial State Information

Yuan Chen, Soummya Kar, and José M. F. Moura

***Abstract*—This paper studies the impact of side initial state information on the detectability of data deception attacks against cyber-physical systems. We assume the attack detector has access to a linear function of the initial system state that cannot be altered by an attacker. First, we provide a necessary and sufficient condition for an attack to be undetectable by any dynamic attack detector under each specific side information pattern. Second, we characterize attacks that can be sustained for arbitrarily long periods without being detected. Third, we define the zero state inducing attack, the only type of attack that remains dynamically undetectable regardless of the side initial state information available to the attack detector. Finally, we design a dynamic attack detector that detects detectable attacks.**

I. Introduction

Cyber-physical systems (CPS) monitor and regulate many critical large-scale infrastructures such as the power grid and water distribution systems. Events such as the Maroochy Shire Council Sewage control incident and the Stuxnet malware attack have brought increased awareness to the issue of securing large scale systems [1], [2]. Smaller applications such as robotic platforms and the modern commercial automobile [3] are also equipped with intercommunicating sensor, computation, and actuator components for a variety of control tasks and can fall suspect to cyber attack. A malicious attacker can hijack the communication channels between the sensor, computation, and actuator components, modify the data values sent between components, and manipulate the system's behavior [4].

To ensure proper operation of CPS, it is necessary to design and implement security measures against attacks. One important aspect of security is attack detection that allows the system to take corrective actions and mitigate damaging behavior. Static attack detectors check the consistency of the system output at a single time step [5], [6], but are unable to detect any attacks on the actuators since they do not consider system dynamics [7]. Reference [7] describes dynamic attack detectors that use the system dynamics, sensing topology, and the history of actuator inputs and sensor outputs to determine whether or not a data deception attack has occurred in a given time window. There are certain attacks, called stealthy or undetectable attacks, that no dynamic detector can detect. Stealthy dynamic attacks change the system output in such a way that the output of the system could arise from the system when it is not under attack [7].

There are several methods to implement attack detection. In [8] and [9], the authors analyze dynamic attacks that go undetected by detectors of bad data (e.g., data resulting from sensor failures) for dynamical systems with process and sensor noise. References [10] and [11] provide algorithms to both detect and reconstruct the dynamic attack. The authors of [12] use sparse optimization techniques to detect and identify deception attacks in electric power systems. Our previous work [13] uses geometric control techniques to analyze the limitations of detecting sparse sensor attacks. A different class of attack detectors, known as active attack detectors, determine the presence of a deception attack by randomly perturbing the system's input and measuring the output [14]. Reference [15] surveys fault detection techniques in dynamic systems that are related to attack detection in CPS. While previous work in attack detection [4], [7], [8], [9], [16] focuses on detectability of attacks, this note precisely clarifies how attack detector performance is sensitive to available information (specifically initial state information) and time horizons.

We present four main contributions. First, we derive a necessary and sufficient condition for an attack to be undetectable when the detector has side initial state information given by an uncorrupted linear function of the initial system state. When the detector has initial state information, an attack is undetectable if and only if it induces a state in the intersection of the system's weakly unobservable subspace and the null space of the side information matrix. Second, we show that an undetectable attack can be maintained if and only if the sum of the change in state produced by the attack and the zero input evolution of the state induced by the attack belong to the system's weakly unobservable subspace. An attack that is undetectable to a certain time point may become detectable at a future time as the detector obtains new sensor measurements. Undetectable attacks that can be maintained indefinitely are a greater security concern than attacks that become detectable after a finite time period. Third, we introduce the zero state inducing attack that is undetectable regardless of the detector's initial state information. We show that such an attack exists if and only if the intersection of the system's output-nulling reachable subspace over one time-step and its weakly unobservable subspace is nonzero. While access to initial state information improves the performance of attack detectors, it is practically important to identify the existence of attacks that are undetectable regardless of the detector's initial state information. Finally, we design a dynamic attack detector that uses side initial state information, has no false alarms, and only misses undetectable attacks.

The rest of this note is organized as follows. In Section II, we specify the system and attack model, review attack detection, introduce side information, and formally state the problem. Section III contains our main technical contributions. Section IV gives the proofs of our main results, section V provides a numerical example illustrating the performance of detectors with side information, and we conclude in Section VI.

II. Background

A. System Model

The cyber-physical system is modeled by

$$\begin{aligned} x(k+1) &= Ax(k) + \overline{B}u(k) + Ba(k), \\ y(k) &= Cx(k) + \overline{D}u(k) + Da(k), \end{aligned} \tag{1}$$

where: $x \in \mathbb{R}^n$ is the system state, $y \in \mathbb{R}^p$ is the system output, $k \in \mathbb{Z}$ is the time index, $u \in \mathbb{R}^m$ is the known input, and $a(k) \in \mathbb{R}^s$ is the unknown attack. Since the input $u(k)$ is known, its contribution to the output $y(k)$ is also known, and therefore, $u(k)$ can be ignored. Thus, for the remainder of the paper, unless otherwise stated, we consider the case of $u(k) \equiv 0$, $\forall k = 0, 1, \ldots$, without loss of generality. Accordingly, we modify the system model to be

$$\begin{aligned} x(k+1) &= Ax(k) + Ba(k), \\ y(k) &= Cx(k) + Da(k). \end{aligned} \tag{2}$$

The matrices B and D describe the capabilities of the attacker. We provide details on the attacker in Section II-C. We use the

Yuan Chen {(412)-268-7103}, Soummya Kar {(412)-268-8962}, and José M.F. Moura {(412)-268-6341, fax: (412)-268-3890} are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15217 {yuanche1, soummyak, moura}@andrew.cmu.edu

This material is based on research sponsored by DARPA under agreement number DARPA FA8750-12-2-0291. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

notation $\Sigma = (A, B, C, D)$ to represent the system[1] in equation (2). Throughout, we make the following assumption.

Assumption 1. *The pair (A, C) is observable.*

Equation (2) with Assumption 1 is a standard model used in the cyber-physical security literature, e.g., [10], [16].

We consider the following sequences: the output sequence (or system output trajectory)

$$Y(T) = \begin{bmatrix} y(0)^T & y(1)^T & \cdots & y(T)^T \end{bmatrix}^T, \tag{3}$$

and the unknown attack sequence

$$E(T) = \begin{bmatrix} a(0)^T & a(1)^T & \cdots & a(T)^T \end{bmatrix}^T, \tag{4}$$

with $T \geq n-1$. An attack occurs when $E(T) \neq 0$. The output trajectory for the deterministic system (1) is

$$Y(T) = \mathcal{O}_T x(0) + \mathcal{M}_T E(T), \tag{5}$$

where $x(0)$ is the system's initial state, $\mathcal{O}_T$ is the extended observability matrix,

$$\mathcal{O}_T = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^T \end{bmatrix}, \tag{6}$$

and $\mathcal{M}_T$ is the input-output matrix,

$$\mathcal{M}_T = \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ CA^{t_1}B & CA^{t_2}B & \cdots & CB & D \end{bmatrix}, \tag{7}$$

where $t_i = T - i$. In our results, we will also work with the extended controllability matrix $\mathcal{C}_T$:

$$\mathcal{C}_T = \begin{bmatrix} A^T B & A^{T-1}B & \cdots & B \end{bmatrix}. \tag{8}$$

The change in state produced by an attack $E(T)$ is $\mathcal{C}_T E(T)$.

We now consider side initial state information. The detector knows the side initial state information

$$y_\Omega = \Omega x(0), \tag{9}$$

where $y_\Omega \in \mathbb{R}^q$ and $\Omega \in \mathbb{R}^{q \times n}$. We call Ω the side information matrix. The matrix Ω having full column rank corresponds to the case in which y_Ω gives full information about $x(0)$, i.e., assuming that we know Ω, we can exactly determine $x(0)$ from y_Ω when Ω is full rank. The matrix Ω being the zero matrix corresponds to the case in which y_Ω gives no information about $x(0)$.

The side information y_Ω captures knowledge of the initial state $x(0)$ from the physical description of the system. For example, consider a remotely controlled vehicle whose state consists of its position and velocity. At $t = 0$ the initial velocity is known to be 0, since, by definition, the system was not running before $t = 0$. We consider the initial position to be unknown since the vehicle is remotely controlled. We emphasize that the side information y_Ω does not rely on sensor measurements. For this reason, the attacker cannot modify the side information y_Ω.

[1]The term "system" refers to the cyber-physical system and attacker collectively. The cyber-physical system gives the A and C matrices of Σ, while the attacker gives the B and D matrices of Σ.

B. Extended System Subspaces

Throughout this note, we use properties of the system's extended observability and reachability subspaces (defined in [17] and [18]) to derive our results. We review their definitions here.

Definition 1 (Weakly Unobservable Subspace $\mathcal{V}(\Sigma)$ [17])**.** *The weakly unobservable subspace of a system Σ, $\mathcal{V}(\Sigma)$, is the subspace of all $x \in \mathbb{R}^n$ such that, for a system with initial condition $x(0) = x$, there exists an input sequence $E(n-1)$ so that the output trajectory is $Y(n-1) = 0$.*

A state $x(0)$ belongs to the weakly unobservable subspace of Σ if and only if there exists an input sequence $E(T)$ such that [17], [18]

$$\mathcal{M}_T E(T) + \mathcal{O}_T x(0) = 0 \text{ for any } T = 0, 1, 2, \ldots$$

References [18], [17], [19], [20] present approaches to calculate a basis for $\mathcal{V}(\Sigma)$.

Another extended system subspace of interest is the output-nulling reachable subspace over k steps.

Definition 2 (Output-nulling Reachable Subspace $\mathcal{W}_k$ [18])**.** *The output-nulling reachable subspace over k steps, $\mathcal{W}_k$, is the subspace of all states $x \in \mathbb{R}^n$ such that there exists an input (attack) sequence $E(k-1)$ that brings the system from $x(0) = 0$ to $x(k) = x$ while producing the output sequence $Y(k-1) = 0$.*

The output-nulling reachable subspace over k steps is the subspace of all states $x \in \mathbb{R}^n$ for which there exists $E(k-1) \in \mathbb{R}^{sk}$ such that $\mathcal{C}_{k-1} E(k-1) = x$ and $\mathcal{M}_{k-1} E(k-1) = 0$.

C. Dynamic Attack Detection: Preliminaries

A dynamic attack detector, ψ, examines the system output $Y(T)$ and side initial state information y_Ω to determine whether or not an attack has occurred:

$$\psi : \mathbb{R}^{p(T+1)} \times \mathbb{R}^q \to \{\text{Attack}, \text{No Attack}\}, \tag{10}$$

where "Attack" means that an attack has occurred. We make the following assumptions.

Assumption 2. *The detector ψ knows the matrices A and C in (2) a priori. The detector ψ does not know the matrices B and D in (2) a priori. The detector ψ a priori does not know $x(0)$ but knows the matrix Ω in (9).*

If we do not impose further restrictions on the detector, then, trivially, we can consider a detector ψ that maps any input to the "Attack" output. For this particular detector, every attack is detectable, but clearly this is not interesting. We restrict our focus to *consistent* attack detectors.

Definition 3 (Consistent Attack Detector [7])**.** *An attack detector ψ is consistent if $\psi(\mathcal{O}_T \theta, \Omega\theta) =$ No Attack for all $\theta \in \mathbb{R}^n$.*

Consistency is a desired property of attack detectors: consistent attack detectors do not produce false alarms. Another desired property of attack detectors is soundness.

Definition 4 (Sound Attack Detector)**.** *A consistent attack detector ψ is sound if $\psi(Y(T), y_\Omega) =$ No Attack for some $Y(T)$ and y_Ω, then, for any other consistent detector $\widetilde{\psi}$, $\widetilde{\psi}(Y(T), y_\Omega) =$ No Attack.*

An sound consistent detector is one that detects all possible attacks without violating the consistency property.

We now provide assumptions on the attacker.

Assumption 3. *The matrix* $\begin{bmatrix} B \\ D \end{bmatrix}$ *is injective*[2].

Assumption 4. *The attacker knows the matrices* A, B, C, D *and* Ω *and the system initial state* $x(0)$ *a priori.*

Assumption 5. *The attacker cannot modify* y_Ω.

Let $E(T)$ be an attack, let $Y(T)$ be the output of the system Σ under attack $E(T)$, and let y_Ω be the side initial state information. Considering only consistent detectors, we define undetectable attacks as follows:

Definition 5 (Undetectable Attack)**.** *An attack* $E(T)$ *is undetectable if, for every consistent detector* ψ *and any* $x(0) \in \mathbb{R}^n$, $\psi\left(Y(T), y_\Omega\right) =$ *No Attack, where* $Y(T) = \mathcal{O}_T x(0) + \mathcal{M}_T E(T)$.

A detectable attack is any attack that is not undetectable. We partition the set of all possible attacks (including $E(T) = 0$), $\mathbb{R}^{s(T+1)}$, into a set of undetectable attacks and a set of detectable attacks.

Definition 6 (Set of Undetectable Attacks $\mathcal{U}^{\Omega,T}$)**.** *The set* $\mathcal{U}^{\Omega,T}$ *is the union of set of all attacks* $E(T) \in \mathcal{R}^{s(T+1)}$ *such that* $E(T)$ *is undetectable and the set that only contains* $E(T) = 0$.

When the system is not under attack (i.e., $E(T) = 0$), consistent detectors report "No Attack", so $0 \in \mathcal{U}^{\Omega,T}$.

Define an extension of an attack as follows:

Definition 7 (Extension of an Attack)**.** *An extension of* $E(T)$, $E(T) \neq 0$, *is an attack of the form*

$$\widehat{E}(T') = \begin{bmatrix} E(T)^T & a(T+1)^T & \cdots & a(T')^T \end{bmatrix}^T, \tag{11}$$

for $T' > T$.

The attack sequence $a(T+1), \ldots, a(T')$ is allowed to be the zero sequence. We provide a necessary and sufficient condition for which an undetectable attack $E(T)$ has undetectable extensions $\widehat{E}(T')$ for all $T' > T$ so that the attack sequence never becomes detectable (even as the attack detector obtains new sensor measurements at each time step). If $E(T)$ does not have an undetectable extension for *all* times $T' > T$, then, at some time $T' > T$, regardless of the attack sequence $a(T+1), \ldots, a(T')$, $\widehat{E}(T')$ is detectable.

Reference [7] provides a necessary and sufficient condition for an attack sequence $E(T)$ to be undetectable when $\Omega = 0$.

Lemma 1 ([7])**.** *The attack* $E(T)$ *is undetectable if and only if*

$$\mathcal{O}_T x(0) + \mathcal{M}_T E(T) = \mathcal{O}_T x'(0)$$

for some initial states $x(0), x'(0) \in \mathbb{R}^n$.

One particular form of attack that is undetectable against systems with no side initial state information is known as the zero dynamics attack.

Definition 8 (Zero Dynamics Attack [4])**.** *A zero dynamics attack is an attack* $E(T) = \begin{bmatrix} a(0)^T & \cdots & a(T)^T \end{bmatrix}^T$ *with*

$$a(k) = \lambda^k g, \tag{12}$$

where $g \neq 0$ *and* $\lambda \in \mathbb{C}$ *satisfy*

$$\begin{bmatrix} \lambda I - A & -B \\ C & D \end{bmatrix} \begin{bmatrix} \theta \\ g \end{bmatrix} = 0. \tag{13}$$

A zero dynamics attack exists if and only if there exists $\lambda \in \mathbb{C}$ for which there is a nonzero solution to (13) [4], [7]. Since, by Assumption 3, the matrix $\begin{bmatrix} B^T & D^T \end{bmatrix}^T$ is injective, and $g \neq 0$, we have that $\theta \neq 0$. By construction, a zero dynamics attack satisfies

$$\mathcal{M}_T E(T) + \mathcal{O}_T \theta = 0.$$

Therefore, a zero dynamics attack satisfies the condition given in Lemma 1, where $\theta = x(0) - x'(0)$. We consider $T \geq n-1$, so $\mathcal{O}_T$ is injective since (A, C) is observable. Since $\theta \neq 0$, a zero dynamics attack produces a nonzero change to the output of the system. Zero dynamics attacks are also related to malicious attacks against distributed function calculation [21].

We introduce the zero state inducing attack:

Definition 9 (Zero State Inducing Attack)**.** *An attack sequence* $E(T)$ *is called a zero state inducing attack if it satisfies* $\mathcal{M}_T E(T) = 0$.

The name zero-state inducing attack refers to the property that such an attack does not change the system sensor output, i.e., the change in output is equal to the response of the system when its initial state is $x(0) = 0$. We show that the zero state inducing attack is undetectable regardless of the detector's side information matrix Ω. It is the only type of attack to remain undetectable even if Ω is full rank.

D. Problem Statement

Consider a system $\Sigma = (A, B, C, D)$ over a time interval $0, 1, \ldots, T$, $T \geq n-1$, with initial state $x(0)$ and side initial state information $y_\Omega = \Omega x(0)$. We consider the following four main problems: 1) find the set of all undetectable attacks, $\mathcal{U}^{\Omega,T}$; 2) determine which attacks $E(T) \in \mathcal{U}^{\Omega,T}$ have undetectable extensions up to any time $T' > T$; 3) determine if there exists an arbitrarily long zero state inducing attack against Σ and; 4) design a consistent detector that uses side information and detects all detectable attacks.

III. Main Results

A. Initial State Information and Undetectable Attacks

First, we find a necessary and sufficient condition for an attack to be undetectable, when the attack detector has side initial state information y_Ω. Let $\mathcal{N}(\Omega)$ be the null space of Ω.

Theorem 1 (Undetectable Attacks with Side Initial State Information)**.** *An attack* $E(T)$ *is undetectable (*$E(T) \in \mathcal{U}^{\Omega,T}$*) if and only if there exists* $\theta \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$ *for which* $\mathcal{M}_T E(T) = -\mathcal{O}_T \theta$.

Theorem 1 states that an attack $E(T)$ is undetectable over the time interval $0, \ldots, T$ if and only if the output contributed by the attack (i.e., $\mathcal{M}_T E(T)$) equals the negative of the output of the system operating without attack from an initial state θ, where θ belongs to the intersection of the system's weakly unobservable subspace, $\mathcal{V}(\Sigma)$, and the null-space of the side information matrix, $\mathcal{N}(\Omega)$. We call θ the state induced by the attack. If $\mathcal{N}(\Omega)$ has dimension strictly less than n (i.e., if the side initial state information is non-trivial), then, by using the side initial state information y_Ω, an attack detector may be able to detect attacks that would otherwise be undetectable (in the absence of side information).

Theorem 1 is valid for any side information matrix Ω.

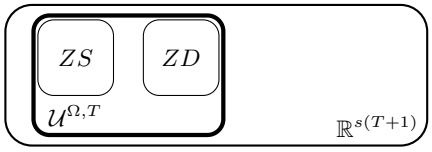
Corollary 1 (No Initial State Information: $\Omega = 0$)**.** *An attack* $E(T)$ *is undetectable if and only if* $\mathcal{M}_T E(T) = -\mathcal{O}_T \theta$ *for some* $\theta \in \mathcal{V}(\Sigma)$ *when* $\Omega = 0$.

By construction, a zero dynamics attack $E(T)$ satisfies $\mathcal{M}_T E(T) + \mathcal{O}_T \theta = 0$, where $\theta \neq 0$ and $g \neq 0$ (which is used to define $E(T)$) is a solution to equation (13). There may be other undetectable attacks aside from zero dynamics attacks when $\Omega = 0$.

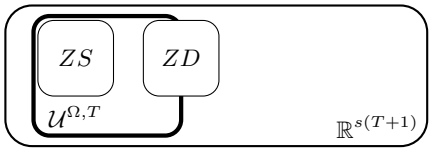
[2]If this matrix is not injective, we can remove the redundant columns to construct an injective matrix. In doing so, we do not change the capabilities of the attacker. Thus, this assumption is made without loss of generality.

Corollary 2 (Full Initial State Information: Ω has full column rank)**.** *An attack $E(T)$ is undetectable if and only if $\mathcal{M}_T E(T) = 0$ when Ω has full column rank.*

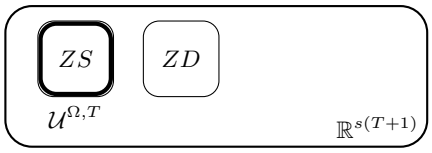
According to Corollary 2, the only type of attack that is undetectable when the initial state is completely known to the detector is the zero state inducing attack. Figure 1 illustrates the results of Theorem 1 and its corollaries. Undetectable attacks presented in the literature [7], [10], [11] rely on the fact that the initial state is unknown to the detector in order to be stealthy. As Theorem 1 and Figure 1 show, however, that even when the detector knows the initial state completely, there may still be undetectable attacks. For the special case of $\Omega = 0$, Theorem 1 is consistent with the results presented in [7].



(a) $\Omega = 0$



(b) $\Omega \neq 0$, Ω is *not* full rank



(c) Ω is full rank

Fig. 1: The set of all undetectable attacks $\mathcal{U}^{\Omega,T}$ depends on the side initial state information available to the attack detector. ZS and ZD are the set of all zero state inducing attacks and the set of all zero dynamics attacks, respectively.

B. Extensions of Undetectable Attacks

Second, we provide a necessary and sufficient condition for an undetectable attack $E(T)$ (with $T \geq n-1$) to have an undetectable extension $\widehat{E}(T')$. Consider an attack $E(T) \in \mathcal{U}^{\Omega,T}$, $E(T) \neq 0$.

Theorem 2 (Extensions of Undetectable Attacks)**.** *There exists an undetectable extension $\widehat{E}(T')$ of $E(T)$ for all $T' > T$ if and only if $\left(\mathcal{C}_T E(T) + A^{T+1}\theta\right) \in \mathcal{V}(\Sigma)$, where θ satisfies $\mathcal{M}_T E(T) = -\mathcal{O}_T \theta$ and $\theta \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$.*

Theorem 2 states that an undetectable attack $E(T)$ has an undetectable extension $\widehat{E}(T')$ for any $T' > T$ if and only if the sum of the change in state produced by the attack ($\mathcal{C}_T E(T)$) and the zero-input state response of the state induced by the attack ($A^{T+1}\theta$) belongs to the system's weakly unobservable subspace ($\mathcal{V}(\Sigma)$). If an attack $E(T)$ satisfies the conditions given in Theorem 2, then for any time $T' > T$, there exists a particular sequence of attacks $a(T+1), \ldots, a(t)$ such that $\widehat{E}(T')$ is undetectable at time T'. Conversely, if an attack $E(T)$ does not satisfy the above condition, then at some time $T' > T$, all extensions $\widehat{E}(T')$ of $E(T)$ are detectable. In this case, all extensions $\widehat{E}(T')$ are detectable by time T' because the detector obtains sensor measurements $y(T+1), \ldots, y(T'+1)$ (even though $E(T)$ was undetectable).

C. Zero State Inducing Attack

Third, we provide a necessary and sufficient condition for the existence of a zero state inducing attack that can be maintained for a arbitrarily long time. We restrict our focus to zero state inducing attacks that begin at time 0. This is to prevent trivial lengthening by appending a fixed length zero state inducing attack $E(T)$ to a zero vector[3].

Theorem 3 (Arbitrarily Long Zero State Inducing Attacks)**.** *There exists an attack $E(T)$ against the system Σ that begins at time 0 such that $\mathcal{M}_T E(T) = 0$ for any $T = 0, 1, \ldots$ if and only if $\mathcal{W}_1 \cap \mathcal{V}(\Sigma) \neq \{0\}$, where $\mathcal{W}_1$ is the output-nulling reachable subspace over one time step.*

Theorem 3 states that there exists an arbitrarily long zero state inducing attack against a system Σ if and only if the intersection of the system's weakly unobservable subspace, $\mathcal{V}(\Sigma)$ and its output-nulling reachable subspace over one step, $\mathcal{W}_1$ is nonzero.

D. Attack Detection With Side Information

We design a consistent dynamic attack detector that detects all attacks $E(T)$ that do not belong to $\mathcal{U}^{\Omega,T}$. Our dynamic detector operates sequentially: at every time instant k (with the exception of an initialization period), the detector collects new sensor outputs $y(k)$ and makes a decision on whether or not the system was attacked in the time period up to time k. Our detector only uses a finite window of sensor measurements in each time interval, which offers advantageous in implementation over detectors that use the entire history of sensor measurements.

First, define $\overline{Y}(k)$ as the l-length window of sensor measurements ending at time k, where $k \geq l-1$:

$$\overline{Y}(k) = \begin{bmatrix} y(k-l+1)^T & y(k-l+2)^T & \cdots & y(k)^T \end{bmatrix}^T. \quad (14)$$

The attack detector makes a decision at every time instant starting at $l-1$. Second, define $\widehat{Y}(k)$, the input to the attack detector at time k, as follows:

$$\widehat{Y}(k) = \begin{cases} \begin{bmatrix} y_\Omega^T & \overline{Y}(k)^T \end{bmatrix}^T, & k = l-1 \\ \overline{Y}(k), & k = l, l+1, \ldots \end{cases}. \quad (15)$$

Third, define the orthogonal projection (operator) onto the range space of a matrix $\mathcal{K}$ (where $\mathcal{K}$ has full column rank) as

$$\Pi_{\mathcal{K}} = \mathcal{K}\left(\mathcal{K}^T\mathcal{K}\right)^{-1}\mathcal{K}^T. \quad (16)$$

We construct the detector ψ as

$$\psi\left(\widehat{Y}(k)\right) = \begin{cases} \text{No Attack}, & \widehat{Y}(k) = \Pi_{\mathcal{K}(k)}\widehat{Y}(k) \\ \text{Attack}, & \text{Otherwise} \end{cases}, \quad (17)$$

where

$$\mathcal{K}(k) = \begin{cases} \begin{bmatrix} \Omega^T & \mathcal{O}_{l-1}^T \end{bmatrix}^T, & k = l-1 \\ \mathcal{O}_{l-1}, & k = l, l+1, \ldots, \end{cases}. \quad (18)$$

The detector decides that no attack has occurred in the time interval $0, \ldots, T$ if $\psi\left(\widehat{Y}(l-1)\right) = \psi\left(\widehat{Y}(l)\right) = \cdots = \psi\left(\widehat{Y}(T)\right) =$ No Attack.

[3]This is not a restriction on the *definition* of the zero state inducing attack. An attack $E(T)$ with nonzero first attack time can still be a zero state inducing attack if $\mathcal{M}_T E(T) = 0$.

Theorem 4 (Consistency and Soundness of ψ)**.** *For $l \geq n+1$, where n is the dimension of the system state space, $\psi\left(\widehat{Y}(l-1)\right) = \psi\left(\widehat{Y}(l)\right) = \cdots = \psi\left(\widehat{Y}(T)\right)$ = No Attack if and only if $Y(T) = \mathcal{O}_T x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$.*

The detector ψ is consistent and sound when the window length l is sufficiently long. The novelty of our detector is its use of the available side information y_Ω. Detectors that do not use side information (e.g., fault detectors such as those presented in [15]) may still detect some attacks, but, following Theorem 1, such detectors may not be sound. That is, there are certain attacks that are only detectable if the detector uses side information y_Ω.

IV. Proof of Main Results

A. Proof of Theorem 1

First, we provide an intermediate result by modifying Lemma 1 to account for attack detectors with side information y_Ω. Consider a system $\Sigma = (A, B, C, D)$ equipped with an attack detector that has side information matrix Ω.

Lemma 2. *An attack $E(T)$ against the system Σ is undetectable if and only if $\mathcal{M}_T E(T) + \mathcal{O}_T x(0) = \mathcal{O}_T x'(0)$ and $\Omega x(0) = \Omega x'(0)$ for some initial states $x(0), x'(0) \in \mathbb{R}^n$.*

We use the above Lemma to prove Theorem 1

Proof (Theorem 1): (If) Let $x(0)$ be the initial state of the system. Let $E(T)$ be an attack such that $\mathcal{M}_T E(T) = -\mathcal{O}_T\theta$ for $\theta \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$. Let $x'(0) = x(0) - \theta$. Then $\mathcal{M}_T E(T) + \mathcal{O}_T x(0) = \mathcal{O}_T x'(0)$. In addition, since $\theta \in \mathcal{N}(\Omega)$, $\Omega x'(0) = \Omega(x(0) - \theta) = \Omega x(0)$. Thus, for any $x(0)$, there exists $x'(0)$ such that $\mathcal{M}_T E(T) + \mathcal{O}_T x(0) = \mathcal{O}_T x'(0)$ and $\Omega x(0) = \Omega x'(0)$, which means, by Lemma 2, $E(T)$ is an undetectable attack. Thus, $E(T) \in \mathcal{U}^{\Omega,T}$.

(Only If) Let $x(0)$ be the initial state of the system. Let $E(T) \in \mathcal{U}^{\Omega,T}$. Then, by Lemma 2, there exists $x'(0) \in \mathbb{R}^n$ such that $\mathcal{M}_T E(T) + \mathcal{O}_T x(0) = \mathcal{O}_T x'(0)$ and $\Omega x(0) = \Omega x'(0)$. Let $\theta = x(0) - x'(0)$. Substituting for θ we have that $\mathcal{M}_T E(T) = -\mathcal{O}_T\theta$ and $\Omega\theta = 0$. Thus, $\mathcal{M}_T E(T) = -\mathcal{O}_T\theta$ for $\theta \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$. ∎

B. Proof of Theorem 2

Proof: (Only If) We show that if there exists an undetectable extension $\widehat{E}(T')$ for all $T' > T$, then, necessarily, $\left(\mathcal{C}_T E(T) + A^{T+1}\theta\right) \in \mathcal{V}(\Sigma)$. Let

$$\widehat{E}(T') = \begin{bmatrix} E(T)^T & a(T+1)^T & \cdots & a(T')^T \end{bmatrix}^T$$

be an undetectable extension of $E(T)$. Since $\widehat{E}(T')$ is undetectable, then, by Theorem 1, it must satisfy $\mathcal{M}_{T'}\widehat{E}(T') + \mathcal{O}_{T'}\theta' = 0$ for some $\theta' \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$.

We first show that $\theta' = \theta$. We partition the matrix $\mathcal{M}_{T'}$ as follows:

$$\mathcal{M}_{T'} = \begin{bmatrix} \mathcal{M}_T & 0 \\ \mathcal{Q}_{T'}^T & \mathcal{M}_{T'-T-1} \end{bmatrix}, \tag{19}$$

where $\mathcal{Q}_{T'}^T = \mathcal{O}_{T'-T-1}\mathcal{C}_T$. Substituting for the partitioned versions of $\mathcal{M}_{T'}$ and partitioning $\mathcal{O}_{T'}$, we have

$$\begin{bmatrix} \mathcal{M}_T & 0 & \mathcal{O}_T \\ \mathcal{Q}_{T'}^T & \mathcal{M}_{T'-T-1} & \mathcal{O}_{T'-T-1}A^{T+1} \end{bmatrix} \begin{bmatrix} \widehat{E}(T') \\ \theta' \end{bmatrix} = 0. \tag{20}$$

From the first block row of equation (20), we have $\mathcal{M}_T E(T) + \mathcal{O}_T\theta' = 0$, and, from the definition of $E(T)$, we have $\mathcal{M}_T E(T) + \mathcal{O}_T\theta = 0$. Thus, $\mathcal{O}_T\theta' = \mathcal{O}_T\theta$. Since $T \geq n-1$ and Σ is observable, $\mathcal{O}_T$ is injective, and $\theta' = \theta$.

Substituting $\theta = \theta'$, the second block row of equation (20) gives

$$\mathcal{O}_{T'-T-1}\left(\mathcal{C}_T E(T) + A^{T+1}\theta\right) + \mathcal{M}_{T'-T-1}\begin{bmatrix} a(T+1) \\ \vdots \\ a(T') \end{bmatrix} = 0. \tag{21}$$

Since there exists an undetectable extension $\widehat{E}(T')$ of $E(T)$ for all $T' > T$, equation (21) must be satisfied for all $T' > T$. In particular, equation (21) is true for $T' = T + n$, which shows that $\left(\mathcal{C}_T E(T) + A^{T+1}\theta\right) \in \mathcal{V}(\Sigma)$.

(If) If $\left(\mathcal{C}_T E(T) + A^{T+1}\theta\right) \in \mathcal{V}(\Sigma)$, then, for all $T' > T$, there exists an attack sequence $\begin{bmatrix} a(T+1)^T & \cdots & a(T')^T \end{bmatrix}^T$ such that equations (21) is satisfied. For all $T' > T$, we construct $\widehat{E}(T')$ by appending $\begin{bmatrix} a(T+1)^T & \cdots & a(T')^T \end{bmatrix}^T$ to $E(T)$. By definition of $E(T)$, we have $\mathcal{M}_T E(T) + \mathcal{O}_T\theta = 0$, where $\theta \in \mathcal{N}(\Omega) \cap \mathcal{V}(\Sigma)$. Combining this fact with equation (21), we see that $\begin{bmatrix} \widehat{E}(T') \\ \theta' \end{bmatrix}$ satisfies equation (20) with $\theta' = \theta$. Thus, we have

$$\mathcal{M}_{T'}\widehat{E}(T') + \mathcal{O}_{T'}\theta = 0,$$

which shows that $\widehat{E}(T')$ is an undetectable extension of $E(T)$. ∎

C. Proof of Theorem 3

Proof: (If) We construct a zero state inducing attack $E(T)$ that begins at time 0 against Σ of arbitrary length T under the condition that $\mathcal{W}_1 \cap \mathcal{V}(\Sigma) \neq \{0\}$. The initial state of the system Σ, $x(0)$, does not affect its extended observability and reachability subspaces, so, without loss of generality, let the system have initial state $x(0) = 0$. If $\mathcal{W}_1 \cap \mathcal{V}(\Sigma) \neq \{0\}$, there exists an attack $a(0) \neq 0$ such that $x(1) = Ba(0)$, $y(0) = Da(0) = 0$, and $x(1) \in \mathcal{V}(\Sigma)$. Since $x(1) \in \mathcal{V}(\Sigma)$, for any T, there exists a sequence of attacks $\begin{bmatrix} a(1)^T & a(2)^T & \cdots & a(T)^T \end{bmatrix}^T$ such that the output $\begin{bmatrix} y(1)^T & y(2)^T & \cdots & y(T)^T \end{bmatrix}^T$ is 0. Thus, for any T, there exists an attack $E(T) = \begin{bmatrix} a(0)^T & a(1)^T & \cdots & a(T)^T \end{bmatrix}^T$ with $a(0) \neq 0$ such that $\mathcal{M}_T E(T) = 0$.

(Only If) We show that if there exists a zero state inducing attack that begins at time 0 for any T against the system Σ, then $\mathcal{W}_1(\Sigma) \cap \mathcal{V}(\Sigma) \neq \{0\}$. Such an attack exists for any T, so it exists for $T = n$. Let

$$E(n) = \begin{bmatrix} a(0)^T & a(1)^T & \cdots & a(n)^T \end{bmatrix}^T$$

be a zero state inducing attack with $a(0) \neq 0$. Since $E(n)$ induces the zero state, we have $\mathcal{M}_n E(n) = 0$, which implies that $Da(0) = 0$. Since $\begin{bmatrix} B \\ D \end{bmatrix}$ is injective and $Da(0) = 0$, we have $x(1) = Ba(0) \neq 0$ and $x(1) \in \mathcal{W}_1$. The sequence

$$\begin{bmatrix} a(1)^T & a(2)^T & \cdots & a(n)^T \end{bmatrix}^T$$

is an input sequence over n steps such that a system with state $x(1) = Ba(0)$ produces zero output over the time period $1, \ldots, n$. Since such an input sequence exists, $x(1) \in \mathcal{V}(\Sigma)$ and $x(1) \in \mathcal{W}_1 \cap \mathcal{V}(\Sigma)$. Since $x(1) \neq 0$, $\mathcal{W}_1 \cap \mathcal{V}(\Sigma) \neq \{0\}$. ∎

D. Proof of Theorem 4

Proof: (If) Let $Y(T) = \mathcal{O}_T x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$. Then, by construction of $\widehat{Y}(k)$,

$$\widehat{Y}(k) = \mathcal{K}(k)A^{k-l+1}x(0). \tag{22}$$

for all $k = l-1, l, \ldots, T$, which means that

$$\Pi_{\mathcal{K}(k)}\widehat{Y}(k) = \widehat{Y}(k), \tag{23}$$

for all $k = l-1, l, \ldots, T$. Thus,

$$\psi\left(\widehat{Y}(l-1)\right) = \psi\left(\widehat{Y}(l)\right) = \cdots = \psi\left(\widehat{Y}(T)\right) = \text{No Attack}.$$

(Only If) We resort to induction.

Base Case: In the base case, we show that if

$$\psi\left(\widehat{Y}(l-1)\right) = \psi\left(\widehat{Y}(l)\right) = \text{No Attack},$$

then $Y(l) = \mathcal{O}_l x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$. Since $\psi\left(\widehat{Y}(l-1)\right) = \text{No Attack}$, we have

$$\widehat{Y}(l-1) = \Pi_{\mathcal{K}(l-1)}\widehat{Y}(l-1), \tag{24}$$

which means that

$$\widehat{Y}(l-1) = \mathcal{K}(l-1)x(0), \tag{25}$$

$$= \begin{bmatrix} \Omega \\ \mathcal{O}_{l-1} \end{bmatrix} x(0), \tag{26}$$

for some $x(0) \in \mathbb{R}^n$. Since $\psi\left(\widehat{Y}(l)\right) = \text{No Attack}$, we have

$$\widehat{Y}(l) = \mathcal{O}_{l-1}x'(0). \tag{27}$$

for some $x'(0) \in \mathbb{R}^n$. From equation (26), we have

$$\begin{bmatrix} y(1)^T & \cdots & y(l-1)^T \end{bmatrix}^T = \mathcal{O}_{l-2}Ax(0), \tag{28}$$

and from equation (27), we have

$$\begin{bmatrix} y(1)^T & \cdots & y(l-1)^T \end{bmatrix}^T = \mathcal{O}_{l-2}x'(0). \tag{29}$$

The pair (A, C) is observable and $l \geq n+1$, so the matrix $\mathcal{O}_{l-2}$ is injective. Thus, combining equations (28) and (29), we have $x'(0) = Ax(0)$. By definition of $\widehat{Y}(l)$ and substituting $x'(0) = Ax(0)$ into equation (27), we have that $y(l) = CA^l x(0)$. Note that $Y(l) = \begin{bmatrix} \overline{Y}(l-1)^T & y(l)^T \end{bmatrix}^T$. Thus, $Y(l) = \mathcal{O}_l x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$.

Induction Step: In the induction step, we assume that if

$$\psi\left(\widehat{Y}(l-1)\right) = \cdots = \psi\left(\widehat{Y}(T-1)\right) = \text{No Attack},$$

then $Y(T-1) = \mathcal{O}_{T-1}x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$. We show that if $\psi\left(\widehat{Y}(T)\right) = \text{No Attack}$ as well, then $Y(T) = \mathcal{O}_T x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$.

Since $\psi\left(\widehat{Y}(T)\right) = \text{No Attack}$, we have

$$\widehat{Y}(T) = \mathcal{O}_{l-1}x'(0), \tag{30}$$

for some $x'(0) \in \mathbb{R}^n$. From the induction hypothesis, we have that $Y(T-1) = \mathcal{O}_{T-1}x(0)$, which means that

$$\begin{bmatrix} y(T-l+1)^T & \cdots & Y(T-1)^T \end{bmatrix}^T = \mathcal{O}_{l-2}A^{T-l+1}x(0). \tag{31}$$

From equation (30), we have

$$\begin{bmatrix} y(T-l+1)^T & \cdots & Y(T-1)^T \end{bmatrix}^T = \mathcal{O}_{l-2}x'(0). \tag{32}$$

The pair (A, C) is observable and $l \geq n+1$, so the matrix $\mathcal{O}_{l-2}$ is injective. As a result, $x'(0) = A^{T-l+1}x(0)$. Substituting $\theta' = A^{T-l+1}$ into equation (32), we have $y(T) = CA^T x(0)$. Note that $Y(T) = \begin{bmatrix} Y(T-1)^T & y(T)^T \end{bmatrix}^T$. Thus, $Y(T) = \mathcal{O}_T x(0)$ and $y_\Omega = \Omega x(0)$ for some $x(0) \in \mathbb{R}^n$. ■

V. Numerical Example

We illustrate our results with an example of a remotely piloted aircraft subject to both nonzero state inducing attacks and zero state inducing attacks. Reference [22] provides a numerical model of the longitudinal dynamics of a remotely piloted aircraft that accounts for the aircraft's physical parameters. We describe the longitudinal dynamics of the aircraft using four state variables: horizontal velocity (x_1), vertical velocity (x_2), pitch rate (x_3), and pitch angle (x_4). The aircraft we consider has two actuators: the elevator (u_1) and the thrust (u_2). The aircraft also has three sensors: the horizontal velocity sensor (y_1), the vertical velocity sensor (y_2), and the pitch angle sensor (y_3).

The evolution of the state variables $x_1, \ldots, x_4$ is determined by physical principles governing the longitudinal flight of the aircraft and depends on physical parameters of the aircraft such as its mass and its pitch moment. The model is linearized about an equilibrium point, so the state variables $x_1, \ldots, x_4$ represent values of the internal states relative to a fixed point (e.g., x_1 in the linearized model is the horizontal velocity of the aircraft relative to an equilibrium horizontal velocity). The linearized, discretized model for the aircraft gives the following dynamics and sensing matrices [22]:

$$A = \begin{bmatrix} 0.992 & 0.030 & -0.003 & -0.977 \\ 0.025 & 0.684 & 1.847 & -0.041 \\ 0.054 & -0.100 & 0.381 & -0.025 \\ 0.003 & -0.006 & 0.068 & 0.999 \end{bmatrix}, \tag{33}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{34}$$

The pair (A, C) in this example is observable.

We consider an attacker modeled by the following B and D matrices:

$$B = \begin{bmatrix} 0.001 & 0.025 & 0 & 0 \\ -3.224 & -0.035 & 0 & 0 \\ -1.995 & -0.021 & 0 & 0 \\ -0.115 & -0.001 & 0 & 0 \end{bmatrix}, \tag{35}$$

$$D = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{36}$$

The attacker can attack both actuators (elevator, u_1, and thrust, u_2) and the horizontal velocity (y_1) and vertical velocity (y_2) sensors. There exists a zero dynamics attack against the system $\Sigma = (A, B, C, D)$.

In this numerical example, we compare the performance of a detector that does not use side information (i.e., the detector's side information matrix is $\Omega = 0$) and the performance of a detector that uses side information matrix

$$\Omega = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}.$$

The detector with nontrivial side information knows the initial horizontal velocity $x_1(0)$. Both detectors are implementations of the windowed detector presented in Section III-D; the only difference between the use of side initial state information.

We construct a zero dynamics attack (as defined in [4] and [7]) against the remotely piloted aircraft. Following equation (12), we construct the zero dynamics attack component wise as

$$a(k) = (10)(.9779)^k \begin{bmatrix} .0324 & 0 & -.6396 & .3007 \end{bmatrix}^T, \tag{37}$$

where $k = 0, \ldots, 30$. The performance of the two detectors are shown in Figure 2. The detector without side information is unable to detect the zero dynamics attack – the detector outputs 0, equivalent to "No Attack" for all times. The detector with side information is

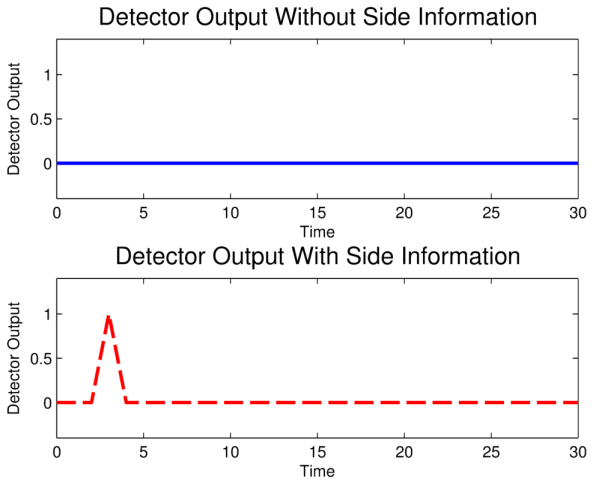


Fig. 2: Detector performance without side information (top) and with side information (bottom) against zero dynamics attack.

able to detect the zero dynamics attack – the detector has an output of 1, equivalent to "Attack" at time t= 3.

VI. Conclusion

In this paper, we studied the effect of side initial state information on the dynamic detection of data deception attacks against cyber-physical systems. First, an undetectable attack induces a state in the intersection of the system's weakly unobservable subspace, $\mathcal{V}(\Sigma)$, and the null space of the side information matrix, $\mathcal{N}(\Omega)$. Second, an undetectable attack $E(T)$ has an undetectable extension to any $T' > T$ if and only if the sum of the change in state produced by the attack, $\mathcal{C}_T E(T)$, and the zero-input state response of the state induced by the attack, $A^{T+1}\theta$, belongs to the system's weakly unobservable subspace, $\mathcal{V}(\Sigma)$. Third, there exists an arbitrarily long zero state inducing attack if and only if the intersection of the system's weakly unobservable subspace, $\mathcal{V}(\Sigma)$, and the system's output-nulling reachable subspace over one step, $\mathcal{W}_1$, is nonzero. Finally, we designed an attack detector that uses side information and detects all attacks that are not undetectable.

References

[1] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proceedings of the 3rd Conference on Hot Topics in Security*, San José, CA, Jul. 2008, pp. 1–6.

[2] A. A. Cárdenas, S. Amin, Z. Lin, Y. H. and. C. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, Hong Kong, Mar. 2011, pp. 355–366.

[3] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *Proceedings of the 2010 IEEE Symposium on Security and Privacy*, Oakland, CA, May 2010, pp. 447–462.

[4] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack models and scenarios for networked control systems," in *Proceedings of the 1st ACM International Conference on High Confidence Networked Systems*, Beijing, China, Apr. 2012, pp. 55–64.

[5] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against power systems in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Chicago, IL, Nov. 2009, pp. 21–32.

[6] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Proceedings of the 2010 IEEE Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2010, pp. 1–6.

[7] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[8] Y. Mo and B. Sinopoli, "Integrity attacks on cyber-physical systems," in *Proceedings of the 1st ACM International Conference on High Confidence Networked Systems*, Beijing, China, Apr. 2012, pp. 47–54.

[9] ——, "False data injection attacks in control systems," in *Proceedings of the 1st Workshop on Secure Control Systems*, Stockholm, Sweden, Apr. 2010, pp. 56–62.

[10] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[11] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attack," *ArXiv e-prints*, Sep. 2013.

[12] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, Mar. 2014.

[13] Y. Chen, S. Kar, and J. M. F. Moura, "Cyber-physical systems: Dynamic sensor attacks and strong observability," in *Proceedings of the 40th International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, Apr. 2015, pp. 1752–1756.

[14] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proceedings of the 47th Annual Allerton Conference*, Monticello, IL, Sep. 2009, pp. 911–918.

[15] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, Nov. 1976.

[16] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Proceedings of the 50th Annual Allerton Conference*, Monticello, IL, Oct. 2012, pp. 1806–1813.

[17] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control Theory for Linear Systems*. Springer, 2001, ch. 7.

[18] B. P. Molinari, "Extended controllability and observability for linear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 136–137, Feb. 1976.

[19] ——, "A strong controllability and observability in linear multivariate control," *IEEE Transactions on Automatic Control*, vol. 21, no. 5, pp. 761–764, Oct. 1976.

[20] D. Rappaport and L. M. Silverman, "Structure and stability of discrete-time optimal systems," *IEEE Transactions on Automatic Control*, vol. 16, no. 3, pp. 227–233, Jun. 1971.

[21] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterations in the presence of malicious agents," in *Proceedings of the 2008 American Control Conference*, Seattle, WA, Jun. 2008, pp. 1350–1355.

[22] R. D. Linehan, K. J. Burnham, and D. J. G. James, "4-Dimensional control of a remotely piloted vehicle," in *Proceedings of the UKACC International Conference on Control '96*, Exeter, UK, Sep. 1996, pp. 770–775.